Accepted Manuscript

Novel Secured Scheme for Blind Audio/Speech Norm-Space Watermarking by Arnold Algorithm

Slami SAADI , Ahmed MERRAD , Ali BENZIANE

PII: S0165-1684(18)30272-X
DOI: https://doi.org/10.1016/j.sigpro.2018.08.011
Reference: SIGPRO 6903

To appear in: *Signal Processing*

Received date: 20 December 2017
Revised date: 16 August 2018
Accepted date: 19 August 2018

Please cite this article as: Slami SAADI , Ahmed MERRAD , Ali BENZIANE , Novel Secured Scheme for Blind Audio/Speech Norm-Space Watermarking by Arnold Algorithm, *Signal Processing* (2018), doi: https://doi.org/10.1016/j.sigpro.2018.08.011

Highlights

- DWT is used after framing the signal
- Sub-sampling into segments for correlation before applying DCT
- Arnold transform is employed to save detection security
- The fully blind detection is accomplished without using the original signal
- Sub-sampling abates robustness against re-sampling attack, increases imperceptibility

Novel Secured Scheme for Blind Audio/Speech Norm-Space Watermarking by Arnold Algorithm

Slami SAADI Ahmed MERRAD Ali BENZIANE

Faculty of Exact Sciences & Computers, Ziane Achour University of Djelfa (UZAD), BP3117, Algeria

Abstract:

In this paper we propose a new scheme for blind watermarking of speech and audio signals. We used the discrete wavelet transform (DWT) after framing the signal, and then we applied the discrete cosine transform (DCT) on each frame. For correlation purpose, sub-sampling is performed to decompose the frame into two segments. For security concern, Arnold transform is employed on the watermark image in order to save detection security. The fully blind detection is accomplished without using the original speech/audio signal and the insertion parameter is not required. Experimental assessment and comparisons with other published schemes demonstrate a good tradeoff between security, capacity, imperceptibility and robustness against various signal processing attacks for both audio and speech signals.

Keywords: Blind watermarking; DWT; DCT; sub-sampling; Arnold Transform; norm space.

1. Introduction

Audio/speech watermarking has many applications such as: copyright protection, usage/Copy tracking, metadata or additional information, multiple data embedding, owner identification, broadcast monitoring and medical applications such as patients reporting. The requirements requested for good watermarking are: imperceptibility which means that the digital watermark should not affect the quality of original audio signal after it is watermarked; robustness that means the embedded watermark data should not be removed or eliminated by unauthorized distributors using common signal processing operations and attacks; capacity that refers to the numbers of bits that can be embedded into the audio signal within a unit of time; and security implying that the watermark can only be detectable by the authorized person.

A semi-blind multiplicative watermarking approach suitable for both audio and speech signals has been presented in [1], in which they used PESQ and PEAQ to optimize the strength factor in order to insert the maximum watermark power while keeping the imperceptibility. A framework jointly exploiting the discrete wavelet packet transform (DWPT) and the discrete cosine transform (DCT) is presented in [2] to perform variable-capacity blind audio watermarking without introducing perceptible distortion and they implemented a neural network for seeking suitable segments for watermark embedding using a perceptual-based quantization index modulation technique. Authors of [3] introduced a flexible variable-dimensional vector modulation (VDVM) scheme to maximize the efficiency of the norm-space DWT-based blind audio watermarking. The watermark embedding is performed in [4] by modulating the vectors in the DCT domain subject to an auditory masking constraint and the abrupt artefacts in frame boundaries are further rectified via linear interpolation over transition areas. Paper [5] presents an adaptive blind audio watermarking algorithm in the wavelet domain to optimize the payload under the perceptual transparency constraints of audio signal by strategically using some of its local features. A blind and robust audio watermarking scheme based on SVD–DCT with the chaotic synchronization code technique is given in [6] by embedding a binary watermark into the high-frequency band of the SVD–DCT block blindly. Lifting wavelet transform (LWT) and singular value decomposition (SVD) are used in [7] by inserting the watermark in the coefficients of the LWT low frequency sub-band taking advantage of both SVD and quantization index modulation (QIM). Authors in [8] outline a package synchronization scheme for blind speech watermarking in the discrete wavelet transform (DWT) domain. Following two-level DWT decomposition, watermark bits and synchronization codes are embedded within selected frames in the second-level approximation and detail sub-bands, respectively where the embedded synchronization code is used for frame alignment and as a location indicator. Using the flexibility of discrete wavelet packet transformation (DWPT) to approximate the critical bands and adaptively determines suitable embedding strengths for carrying out quantization index modulation (QIM), an audio blind watermarking scheme is presented in [9]. In order to protect the digital audio and video products copyright in the network, an improved audio blind watermarking algorithm scheme based on DWT and SVD is proposed in [10]. A new secured chaotic audio watermarking scheme based on self-adaptive particle swarm optimization (SAPSO) and quaternion wavelet transform (QWT) is suggested in [11]. Combining the robustness of vector norm with that of the approximation components after the DWT, a blind and adaptive audio watermarking algorithm is given in [12], where a binary image encrypted by Arnold transform as watermark is embedded in the vector norm of the segmented approximation components. Authors in [13] used the LWT and QR decomposition for audio copyright protection in which, the watermark information is embedded into the largest element of the upper triangular matrix obtained from the low frequency LWT coefficients of each frame.

The feature coefficients cross-correlation degree of speech signal is defined, and the property is discussed, which demonstrates that the feature is very robust in [14]. Then a new watermark embedding method based on the feature is explored, aiming to enlarge the embedding capacity and to solve the security issue of watermark schemes based on public features. In [15], a new blind audio watermarking scheme based on SVD using Angle-Quantization is suggested by embedding the watermark into the angle between the largest singular value and second largest singular value of each diagonal matrix by quantization. Authors in [16] present a secure, robust, and blind adaptive audio watermarking algorithm based on SVD in the DWT domain using synchronization code.

In our proposed scheme, various combinations are used based on DWT and DCT, appending decomposing technique called sub-sampling which it used for watermarking images in [17] and embedding in the norm space, which is a numerical analysis of the linear algebra and can improve the robustness of the algorithm, because the watermark embedded in the norm can be spread throughout all the samples [12]. We also used Arnold transform to encrypt our watermark and grantee the security.

2. Discrete Wavelet Transform (DWT)

The DWT is a novel transform that gives a time-frequency representation of a signal [10]. It was developed to overcome the small variations of the signal with time that are not well covered by Fourier transform in frequency domain. It can as well be practical to analyze non stationary signals [10]. And it is used in a large scale for signal processing purposes [18-19]. DWT decomposes an input signal ***S*** into two sets of coefficients, at the heart of DWT is a pair of filters: low pass and high pass, the approximation coefficients cA1 (low frequencies) are produced by passing the signal throughout low pass filter, the details coefficients cD1 (high frequencies) are produced by passing the signal throughout high pass filter, followed by downsampling.

Depending on the purpose, the signal is decomposed on multi-level discrete wavelets [20], where the next decomposition level splits the approximation coefficients cA1 in two parts using the same scheme, replacing S by cA1, and producing cA2 and cD2. Fig.1 illustrates 2 phases DWT decomposition:

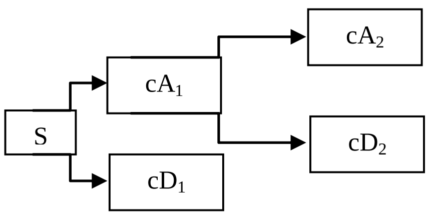


Figure.1. 2-levels DWT decomposition

Inverse DWT process reconstructs or synthesizes the original signal by assembling those components back without loss of information [21], the up-sampling operator is used to recompose the samples eliminated by down sampling. Fig.2:

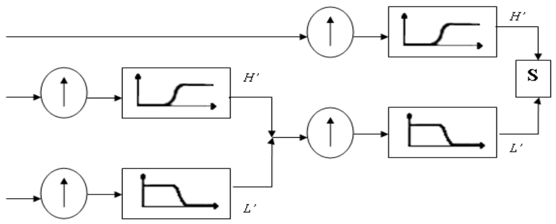


Figure.2 rebuilding a decomposed signal with IDWT

In our case, we use the *Haar* wavelet function (filter) implemented in MATLAB to obtain approximation coefficients which are less sensitive for the human auditory system when we embed the watermark image. In our case one level decomposition is enough.

3. Discrete Cosine Transform (DCT)

The DCT is a recognized transform capable to illustrate fragments of an audio signal in terms of summing up of cosine functions in diverse frequencies. One of the major important obvious features of DCT transform is energy storage in a small number of samples. This feature is used to decrease curvature of the original signal in speech watermarking process [22-23]. The discrete cosine transform is a scheme for converting a signal into fundamental frequency components. The DCT definition of a 1-D sequence of length N is:

$$c(u) = a(u)\sum_{x=0}^{N-1} f(x)\cos\left(\frac{\pi(2x+1)u}{2N}\right), \qquad (1)$$

$$\text{For } u = 0,1,2,\dots,N-1$$

Where, *x(n)* is the original speech signal and N is the number of samples.

In analogous way, the inverse transform is expressed as:

$$f(x) = \sum_{x=0}^{N-1} a(u)\mathrm{c(u)}\cos\left(\frac{\pi(2x+1)u}{2N}\right), \qquad (2)$$

$$\text{For } u = 0,1,2,\dots,N-1$$

In both equations, *a(u)* is defined as:

$$a(u) = \begin{cases} \frac{1}{\sqrt{N}} & u = 0, \\ \sqrt{\frac{2}{N}} & u \neq 0. \end{cases} \qquad (3)$$

The characteristics of this algorithm are strong, well hidden and resistant to a variety of signal deformation resistance. The digital watermark in the DCT transform domain has important ability of lossy compression resistance. The disadvantage is its immense amount of calculations [24].

4. Blind and non-blind watermarking

Blind watermarking does not need the host signal for watermark recognition. On the different, digital watermarking that necessitates the host signal to take out the watermark is non-blind. In general, watermark detection is further robust if the original un-watermarked data are accessible. Though, admission to the original host signal cannot be justified on the whole real-world situations. Then, blind watermarking is further flexible and useful [25].

In several applications, the recognition algorithm can employ the original audio signal to take out watermark from the watermarked signal (informed detection) [26]. It regularly considerably gets better the detector performance; since the watermark information is taken out through deduct the original signal from the watermarked signal. Though, if the detection algorithm does not have admission to the original signal (blind recognition) and this incapacity significantly reduces the quantity of information that can be buried in the original signal. The entire procedure of embedding and extracting of the watermark is modeled as a communication channel where watermark is deformed due to the existence of strong interference as well as channel effects.

5. Arnold Scrambling Transform

The KxK binary watermark image W is transformed into W' by Arnold transformation to reduce the autocorrelation coefficient of image and next the privacy of watermark is reinforce [27]. Arnold transformation is cyclic and while it is iterated occasionally the original signal will be reached. The Arnold scrambling algorithm [28] has the characteristic of ease and periodicity, so it is used usually to offer an extra level of safety all along through digital watermarking. Arnold Transform is well recognized as cat look transforms and is just appropriate for N × N dimension signals. It is defined as:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} mod\ N. \qquad (4)$$

where (x, y) are the coordinates of original watermark and (x', y') are the coordinates of scrambled watermark. N is the height or size of the signal which is to be processed and *mod N* is modulo N (Euclidian division rest). Arnold Transform is periodic in nature. The decryption of signal depends on the scrambling key which can be employed as secret key and defines the number of times it has been scrambled.

6. Proposed scheme

Under watermarking terms, the watermark bits must be distributed along the whole speech/audio signal, and for that we decomposed the signal into many segments equal to the number of bits we want to embed, then we apply DWT to extract the approximation coefficients and put the watermark bits there, where the human auditory system is less sensitive. It allowed us making the watermark strong and inaudible with keeping the imperceptibility. And we also applied DCT in order to obtain two vectors having convergent values following it by sub-sampling decomposition into frames for correlation purpose. This decomposition abates a little robustness against the re-sampling attack but gives our proposed design other advantages against other attacks and allows the imperceptibility to remain very high. Extraction is blind in our proposed design, without using original signal. The decomposed speech/audio signal into segments is subjected again to DWT and DCT transforms, then the produced vectors are sub-sampled and normalized before extracting the bits used to construct the image and apply the inverse of Arnold transform using the key used in the embedding process to produce the watermark image (Arnold transform is employed to increase security). The steps below explain more the two processes: embedding and extraction:

Embedding process:

Step 1: Insert watermark image $\mathbf{WI}_{\mathbf{NxN}}$

Step 2: For the input speech/audio signal **x** decomposed into **N×N** segments;

Step 3: Scramble watermark image $\mathbf{WI}_{\mathbf{NxN}}$ by Arnold transform using a key and restructure into one dimensional;

W={w (j),1≤ j≤J}, where J=NxN;

For each frame (F_j,1≤ j≤ NxN**) apply the steps (4~12)**

Step 4: Apply 1-level DWT with **'db1'** produces **cA1** and **cD1**

cA: represents the low frequencies (approximation coefficients); **cD**: represents the high frequencies (detail coefficients);

Step 5: apply DCT on **cA1** produces vector named **V**;

Step 6: decompose the vector **V** into two (correlated) sub-vectors $\mathbf{V_1}$ and $\mathbf{V_2}$ using the following sub-sampling operations:

$$\mathbf{V1}\,(\mathrm{k}) = \mathbf{V}(2\mathrm{k}), \qquad (5)$$

$$\mathbf{V2}\,(\mathrm{k}) = \mathbf{V}(2\mathrm{k}-1). \qquad (6)$$

Where k=1,., length of V/2.

Step 7: apply the norm of $\mathbf{V_1}$ and $\mathbf{V_2}$ produces $\mathbf{nrm_{V1}}$ and $\mathbf{nrm_{V2}}$ respectively as the following formulas:

$$\begin{cases} \mathbf{nrm_{V1}} = \sigma_1 = \|V_1\| = \sqrt{\sum_{i=1}^{n} V(i)_1^2}\,, & (7) \\ \boldsymbol{u_1} = \dfrac{\boldsymbol{V_1^t}}{\|\boldsymbol{V_1}\|} = \dfrac{\boldsymbol{V_1^t}}{\boldsymbol{\sigma_1}}\,, & (8) \end{cases}$$

$$\begin{cases} \mathbf{nrm_{V2}} = \sigma_2 = \|V_2\| = \sqrt{\sum_{i=1}^{n} V(i)_2^2}\,, & (9) \\ \boldsymbol{u_2} = \frac{\boldsymbol{V_2^t}}{\|\boldsymbol{V_2}\|} = \frac{\boldsymbol{V_2^t}}{\boldsymbol{\sigma_2}}\,. & (10) \end{cases}$$

$\mathbf{V_1}$, $\mathbf{V_2}$, $\mathbf{u_1}$ and $\mathbf{u_2}$ are a 1 × n vectors, $\boldsymbol{\sigma_1}$ and $\boldsymbol{\sigma_2}$ are the norm of $\mathbf{V_1}$ and $\mathbf{V_2}$ respectively

Step 8: Embedding the bit

$$\mathbf{nrm} = \frac{\mathbf{nrm_{V1}} + \mathbf{nrm_{V2}}}{2}\,, \qquad (11)$$

If (**W(j)=**1)

$$\begin{cases} \mathbf{nrm_{V1}} = \text{nrm} + \Delta; & (12) \\ \mathbf{nrm_{V2}} = \text{nrm} - \Delta; & (13) \end{cases}$$

Else

$$\begin{cases} \mathbf{nrm_{V1}} = \text{nrm} - \Delta; & (14) \\ \mathbf{nrm_{V2}} = \text{nrm} + \Delta; & (15) \end{cases}$$

End

Step 9: Construct V'$_1$ and $\mathbf{V'_2}$ with modified norm of each segment as these formula:

$$\boldsymbol{V'_1} = \mathbf{nrm_{V1}} u_1^t\,; \qquad (16)$$

$$\boldsymbol{V'_2} = \mathbf{nrm_{V2}} u_2^t; \qquad (17)$$

Where $\mathbf{u_1}$ and $\mathbf{u_2}$ calculated on the step 7

Step 10: Combine the two sub-vectors **V'**$_1$ and **V'**$_2$ using the opposite operation in step 6 produce the vector **V'**:

$$\mathbf{V}'\,(2k) = \mathbf{V}'1\,(k); \qquad (18)$$

$$\mathbf{V}'\,(2k-1) = \mathbf{V}'2\,(k); \qquad (19)$$

Where k=1,..., length of V/2

Step 11: Apply **IDCT** on the modified vector **V'** produces modified approximation **cA1'**;

Step 12: Apply **IDWT** on **cA1'** and **cD1** produces modified frame;

Step 13: Reconstruct the watermarked speech/audio signal with modified frames.

Extraction process:

Step 1: For the input speech/audio signal **x'** decomposed into **N×N** segments;

For each frame (F_j,1≤ j≤ NxN**)**

Step 1: Apply steps (4~7) of the embedding process

Step 2: Extraction of the bit

If ($\mathbf{nrm_{V1}}$> $\mathbf{nrm_{V2}}$**)**

$$W(j) = 1; \qquad (20)$$

Else

$$W(j) = 0; \qquad (21)$$

End

Step 3: Construct the image with extracted bits

Step 4: Apply inverse of Arnold transform using key used in the embedding process to produce the watermark image

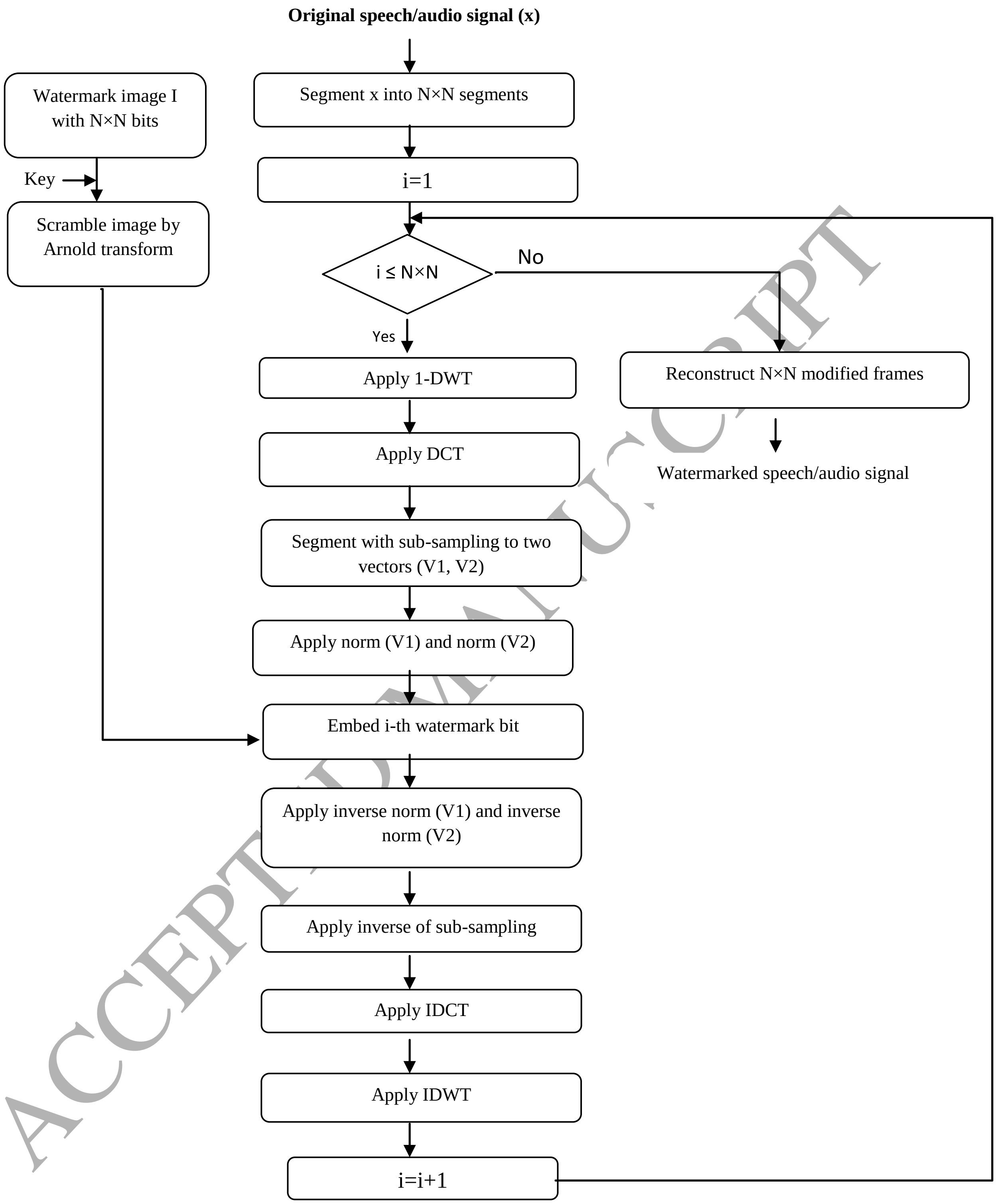


Figure.3 : Watermark Embedding Process

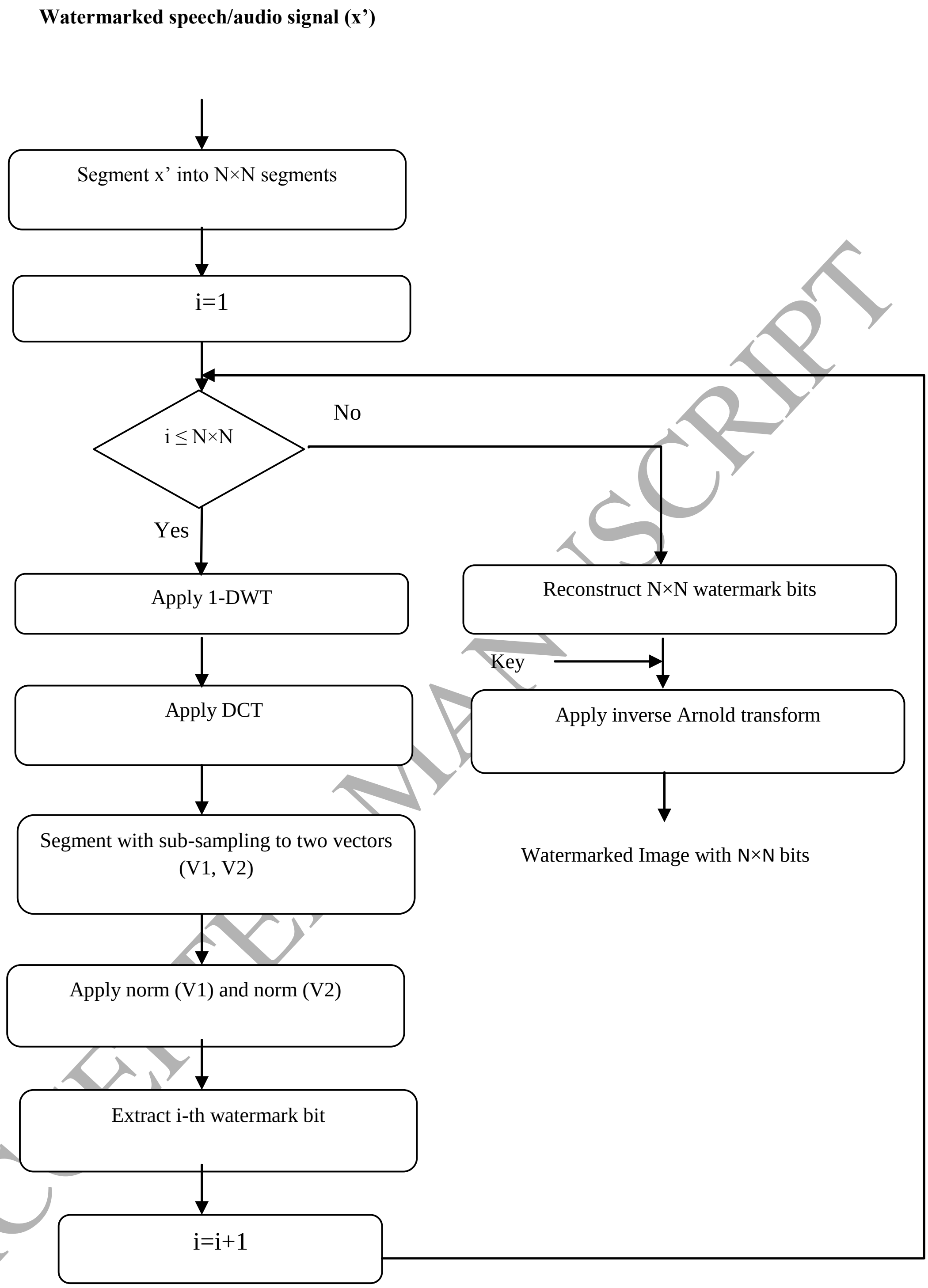


Figure.4: Watermark Extracting Process

7. Experimental results

This section presents all results. All simulations are implemented on Windows PC having Intel 2.2GHz processor and 2GB RAM. All the experiments are performed using MATLAB 7.10.0 on different speech/audio signals which are stored as 16 bit mono wave file, and frequency 44100 Hz.

In order to evaluate the performance of the proposed scheme in real conditions, simulations are performed on different lengths of speech/audio signals included and also different types of human speech signals (male and female) and different languages (English and French).

All of the speeches are downloaded from reference [29], SQAM file (Sound Quality Assessment Material) recording for subjective tests. We edit the speech/audio file to change stereo to mono and we use two binary images as watermarks (UZAD image which it used in all experiments and star image which it used only in the experiments results in tables 5,6 and 7), Fig.5, Fig.6 show them, respectively :

a) Original b) Scrambled image

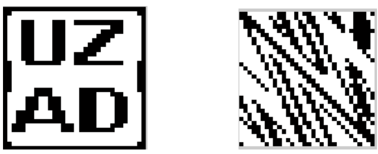


Figure.5: watermark image (UZAD)

a) Original b) Scrambled image

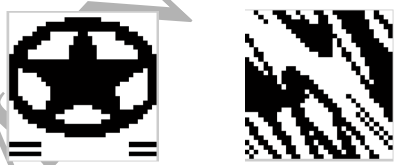


Figure.6: watermark image(STAR)

7.1. Imperceptibility:

Imperceptibility or inaudibility means that watermark embedded into the host signal is inaudible; in this simulation as the majority of this work we use various measurements to assess the quality of the watermarked speech/audio signal. The first is signal-to-noise ratio (SNR) [4] defined as:

$$SNR = 10\log\left(\frac{\sum_{a=1}^{M} S^2(a)}{\sum_{a=1}^{M}\big(S(a) - S'(a)\big)^2}\right), \qquad (22)$$

The second is the Segmental Signal-to-Noise Ratio (SSNR) [30] which is an improvement with respect to conventional SNR measure and it was created to handle the dynamic nature of non-stationary signals such as speech. The definition of SSNR is:

$$SSNR = \frac{1}{N}\sum_{m=1}^{N} SNR_m. \qquad (23)$$

N is the number of frames in the signal

The SNR does not take into account the specific characteristics of the human auditory system, but it can just give a general idea of imperceptibility [31]. Thus, we also employed one of the most popular methods called mean opinion score (MOS) [6,7,31,32] which conducts to provide a better test of inaudibility based on human perception. Ten listeners participated in the practical test and asked to classify the difference between the original and the watermarked speech/audio in terms of 5-points Mean Opinion Score (MOS) with impairment scale defined in Table 1 [31]. To measure the quality of the proposed speech/audio signal, we averaged values of all participants.

Table 1: MOS grading scale

MOS	Description
5	Imperceptible
4	perceptible but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Table 2: SNR, SSNR and MOS of Speech type signal

Speech	**SNR**	**SSNR**	**MOS**
spme50_1	29,7432	35,2420	4,4
spmf52_1	30,2990	35,1564	4,6
spfe49_1	30,5078	35,4074	4,6
average	30,1833	35,2686	4,53

Table 3: SNR, SSNR and MOS of Audio type signal

Audio	**SNR**	**SSNR**	**MOS**
bass47_1	30.0425	35.5654	4,7
gspi35_2	32.1148	33.5571	4,8
average	31,0786	34,5612	4,75

Tables 2 and 3 show values of different measurements for different speech/audio signals results from our proposed method (DWT, DCT, Sub-Sampling, Norm Space, Arnold), so it is clear that the SNR satisfy the requirement of International Federation of the Phonographic Industry (IFPI) with the SNR above 20 db, and it can be up to 30 db which means that our proposed scheme can get better perceptual quality than the previous methods. In addition, we can see that the SSNR is greater than the SNR which means that there is no camouflage.

However, the values of MOS resulting from our proposed method are high, which indicates that the watermarked speech and audio signals are perceptually indistinguishable from the original ones.

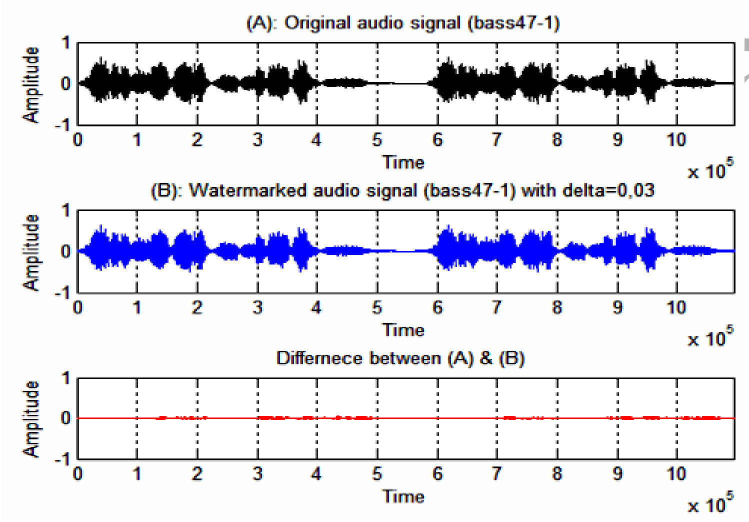


Figure.7: Waveforms of the original and watermarked audio (bass47_1) and difference between them

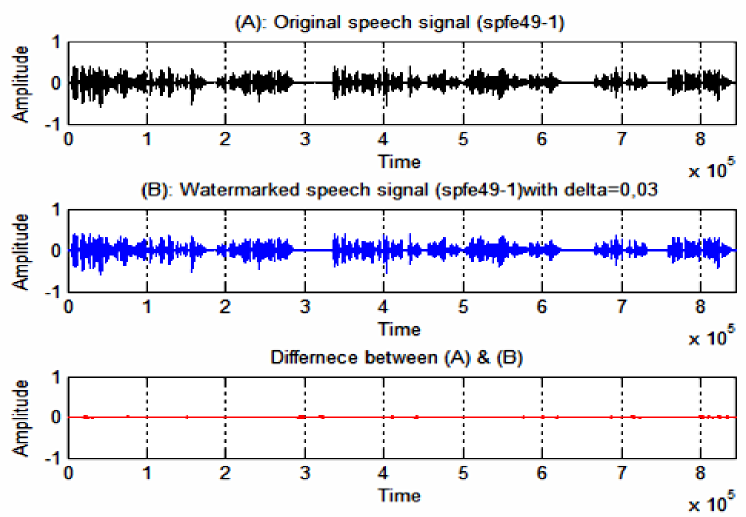


Figure.8: Waveforms of the original and watermarked speech (spfe49_1) and difference between them

Fig.7 illustrates the time waveforms of the original and watermarked audio signal and differences between them respectively, which present the inaudibility by our algorithm. It can be seen that there is only a little visual difference which indicates that our algorithm possesses good transparency.

By observing the waveforms in Fig.8 of the original speech signal (A) and the watermarked version (B) and the difference between them, we can conclude that there is almost no difference.

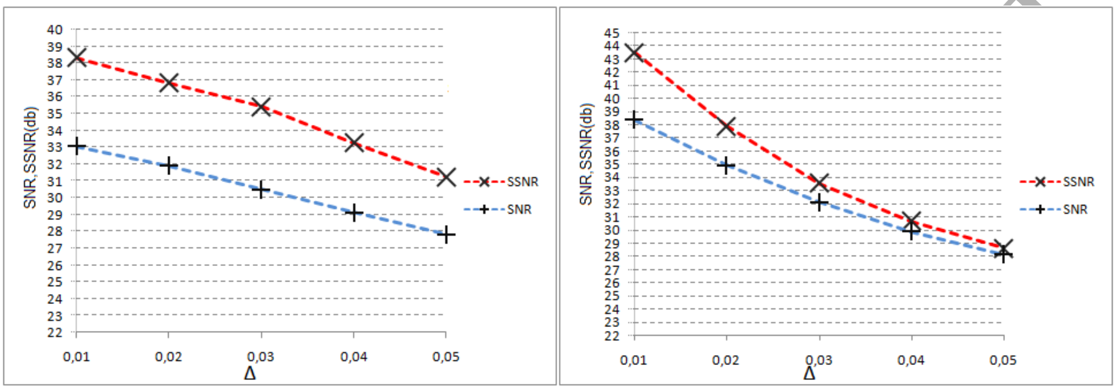


Figure.9: SNR and SSNR versus the Δ for audio and speech signal (on the left: spfe49_1 speech and on the right: gspi35_2 audio)

Fig.9 shows the SNR and SSNR versus the Δ (quantization step) for audio and speech signal (the left: spfe49_1 speech and on the right: gspi35_2 audio). As seen, whenever Δ increases, SNR and SSNR decrease. This is because the norm values are far from their original state (where the bits are embedded), and thus there are a distortions in the original speech/audio signals. Also we can observe that the values of SSNR didn't come down inferior the values of SNR and always stay on up which indicates that there is no camouflage using the process of embedding the watermark.

Robustness:

Robustness is a measure of the watermark against attempts to eliminate or corrupt it, intentionally or accidentally, by different kinds of digital signal processing attacks. For the evaluation of robustness, this simulation examines the bit error rates (BER) between the original watermarking image and the extracted watermarking image. BER is defined by the following expression [32]:

$$BER = \frac{B_{ERR}}{N} \times 100\% \,, \qquad (24)$$

Where B_{ERR} is the number of erroneous bits and N is the total number of bits

Zero means that the attack doesn't have any effect on the watermark and the extraction is successful. Also we employed normalized correlation coefficient (NC) which expresses the similarity between extracted watermarking image and original watermarking image after being attacked and it is defined by the following expression [33]:

$$\mathrm{NC}(w, w') = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N} w(i,j)w'(i,j)}{\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w^2(i,j)}\sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} w'^2(i,j)}}, \qquad (25)$$

where NxN is the size of watermark. *W(i,j)* and *W'(i,j)* are the watermark and recovered watermark images, respectively. One is the best value for NC and it shows that the inserted watermark is extracted successfully.

In order to test the robustness of the proposed algorithm, separately we attack the watermarked version using typical signal processing manipulations

AWGN: Add white Gaussian noise to the vector watermarked speech/audio signal, measuring the power of the audio-speech before adding noise.

Re-sampling: The watermarked speech/audio was down-sampled to half the original sampling rate and then up-sampled back to the original sampling rate.

Re-quantization: 16 bits per sample watermarked speech/audio signals is quantized down to 8 bits per sample.

Echo: We add an echo signal with a different delay and decay of to the watermarked speech/audio signal.

Amplification: The amplitude of the watermarked speech/audio signal is rescaled by ±10%,±15% and ±20%

Cropping: We set the number of samples of the watermarked speech/audio signal to zero randomly.

Table 4: results of robustness against different type of signal processing attacks for audio signal (bass47_1)

The attacks		Watermark images					
		UZAD			STAR		
		SNR between WAS and AWAS	BER %	NC	SNR between WAS and AWAS	BER %	NC
Without attacks		Inf	00	1	Inf	00	1
AWGN		18.0719	00	1	18.0062	00	1
Echo (0.13,0.33)		17.5284	00	1	17.4828	00	1
Re-sampling		40.7000	5.0781	0.9595	41.6001	4.5898	0.9544
Re-quantizaton		31.5877	00	1	31.5842	00	1
Cropping (10000)		20.3075	00	1	20.2556	00	1
Amplification	+20%	19.8671	00	1	20.7071	00	1
	-20%	20.7070	00	1	19.8670	00	1

Table 5: results of robustness against different type of signal processing attacks for speech signal (spme50_1)

The attacks		Watermark images					
		UZAD			STAR		
		SNR between WAS and AWAS	BER %	NC	SNR between WAS and AWAS	BER %	NC
Without attacks		Inf	00	1	Inf	00	1
AWGN		18.0519	00	1	18.0042	00	1
Echo (0.15, 0.32)		12.1942	00	1	12.2625	00	1
Resampling		34.7870	4.9805	0.9603	35.0483	4.4922	0.9554
Re-quantizaton		31.5584	00	1	31.5546	00	1
Cropping (10000)		19.0726	00	1	18.9774	00	1
Amplification	+20%	21.2669	00	1	21.9869	00	1
	-20%	21.9868	00	1	21.2668	00	1

Table 6: results of robustness against different type of signal processing attacks for speech signal (spmf52_1)

The attacks		Watermark images					
		UZAD			STAR		
		SNR between WAS and AWAS	BER %	NC	SNR between WAS and AWAS	BER %	NC
Without attacks		Inf	00	1	Inf	00	1
AWGN		18.0614	00	1	18.0050	00	1
Echo (0.12, 0.3)		16.1849	00	1	16.6254	00	1
Resampling		30.2428	4.9805	0.9603	30.3407	4.4922	0.9554
Re-quantizaton		32.0982	00	1	32.0967	00	1
Cropping (3000)		24.9535	00	1	24.6027	00	1
Amplification	+20%	22.6162	00	1	23.2362	00	1
	-20%	23.2361	00	1	22.6161	00	1

Table 4, Table 5 and Table 6 show the robustness of our proposed method using different audio and speech signals (bass47_1, spme50_1 and spmf52_1) without attack and with various attacks. The low SNR between watermarked speech/audio signal (WSS/WAS) and attacked watermarked speech/audio signal (AWSS/AWAS) demonstrates that the majority of attacks used for evaluation of the robustness were very strong such as: AWGN, adding Echo, cropping and amplification attacks. However the majority of the BER values are zeros and the majority of NCs values are ones which means that the process of detection can detect the inserted watermark successfully. It indicates that the watermark system adopted has good robustness performances. So that all attacks can't degrade the watermark except in re-sampling attack, but that's not a problem because the BER is low in this situation and we can still identify our watermark.

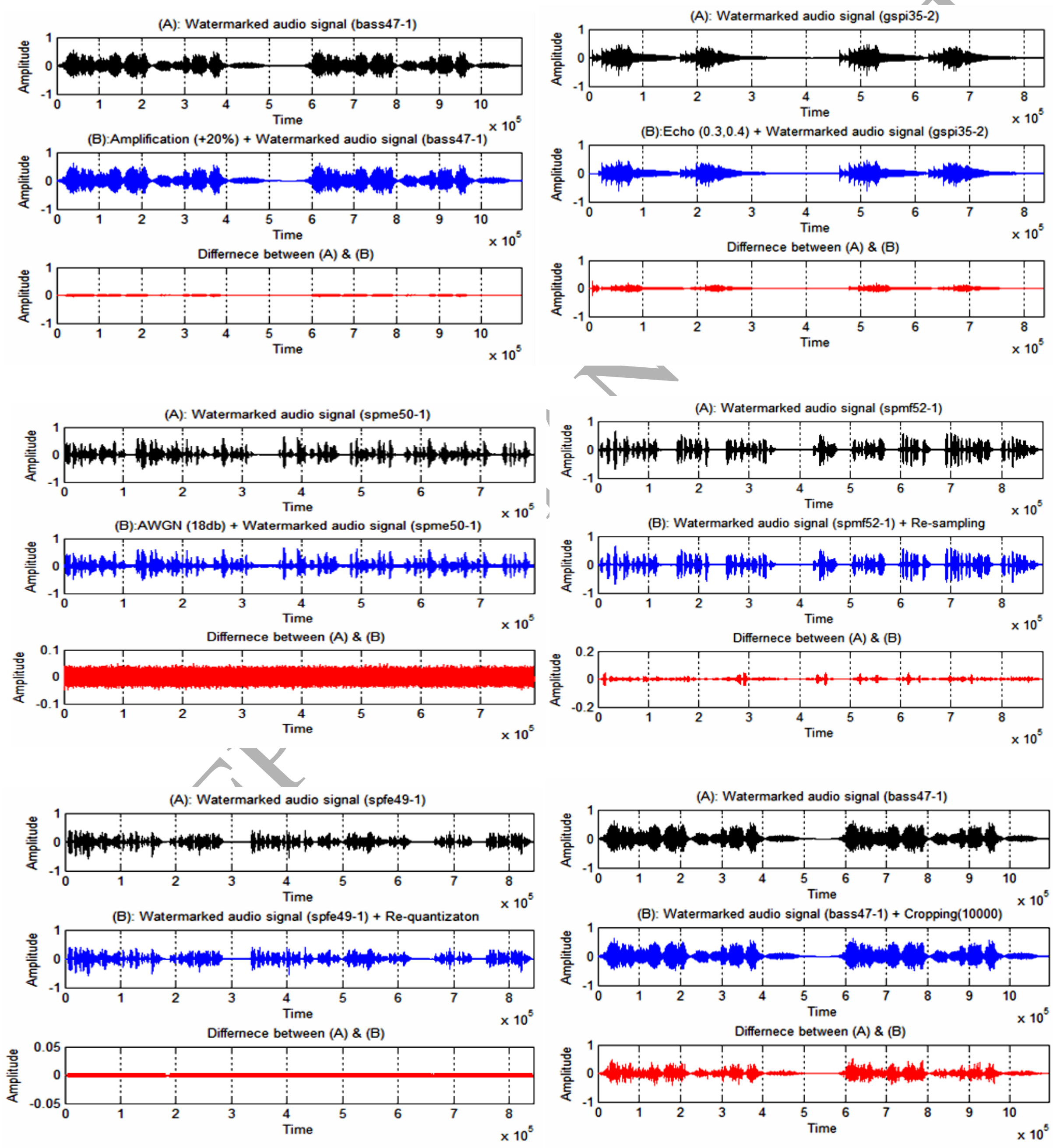


Figure.10: the used different attacks and their effects on original watermarked signals

In Fig.10, we can observe that the attacks used are very strong and effects on the signal. This figure explains more the strong attacks used so that there exists a little difference by the attacks: re-quantization and re-sampling. The difference is noticeable in the attack of amplification and AWGN. Big differences are observed in the echo and cropping attacks between watermarked speech/audio signal and the attacked speech/audio signal.

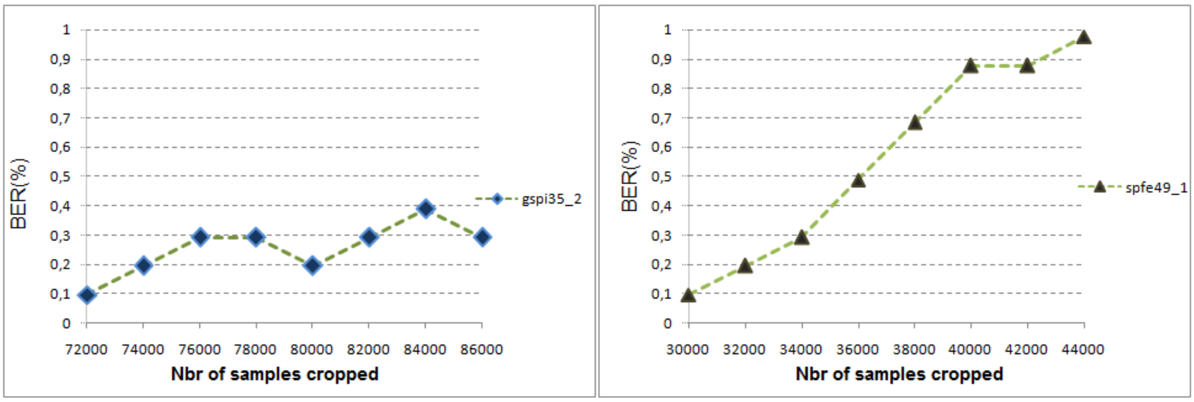


Figure.11: BER vs cropping for audio-speech signal (on the left gspi35_2 audio, on the right spfe49_1 speech)

Fig.11 illustrates the BER values versus increasing number of samples that are cropped in the audio and speech signals. BER remains small under 1% although thousands of samples were set as zero randomly. Although the cropping was changed by 14 thousands cropped samples, the BER remains small and did not exceed 1%.

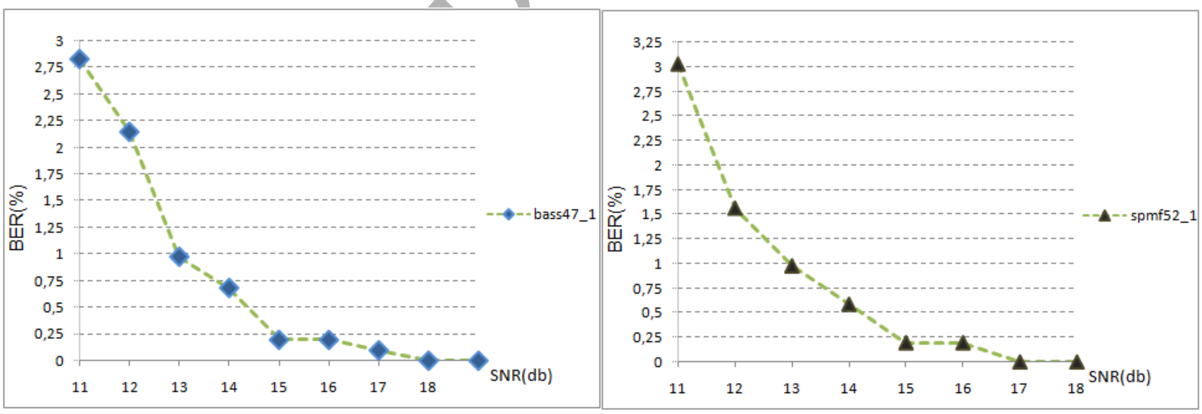


Figure.12: BERs vs AWGN attacks for audio-speech signal (on the left bass47_1 audio, on the right spmf52_1 speech)

Fig.12 shows the BER after different SNR of AWGN attacks. Although all of these attacks are strong and influential on the signal significantly, BER is small at SNR=11db (<3%) and null at SNR=18db. This confirms the robustness of the watermark inserted in speech/audio signal. The lower the strength of AWGN SNR, the more obvious is the watermark.

7.2. Data payload

Data payload is defined as the number of bits embedded in a one second audio fraction [25], and is measured in bits per second (bps). Suppose that S is the duration of the original audio signal in seconds and K is the number of embedded watermark bits, the capacity of the proposed scheme C is expressed as follows [34]:

$$C = \frac{K}{S}\ bps\,. \qquad (26)$$

Table 7: capacity measures for different audio and speech signals

Audio/Speech	bass47_1	gspi35_2	spme50_1	spmf52_1	spfe49_1
capacity	41.19	53.87	57.03	51.17	53.36

The capacity in Table 7 is not too high but it is sufficient as the conditions of IFBI are set to 20b/s a satisfied because the goal is reached, the watermarking is very robust and high imperceptibility is attained.

8. Comparisons

From the comparison results in Table 8, we can see that our proposed (DWT, DCT, Sub-sampling, Norm-space, Arnold) scheme can obtain a relatively high imperceptibility and good payloads results, since SNR and MOS results are higher than almost all other published methods selected for comparison. It demonstrates the preference for our scheme. Besides, the payload in our scheme is lower than in [15] and [8], but it is relatively high compared to the other selected methods.

Table 8: summary of comparisons with seven methods cited in literature

Methods	Average of SNR (db)	Capacity b/s	Type	Average of MOS
DWT-SVD in [10]	20,7	27,56	Speech	4,4
	21,2		Audio	4,65
SVD-AQ in [15]	30,3	172,39	Audio	-
DWT-AMM in [8]	21,932	200	Speech	3,25
CCCD in [14]	25,777	49	Speech	-
DWPT-Multiplication in [35]	28,08	31,25 -125	Speech	4,11
Adaptive DWT SVD in [16]	24,37	45,9	Audio	4,46
Method in [25]	30,0675	17,2	Audio	-
Our proposed scheme (DWT- DCT- Sub sampling - Norm space – Arnold)	**31,0786**	**41.19-53.87**	Audio	4,53
	30,1833	**51.17-57.03**	Speech	4,75

Table 9: comparisons between our proposed scheme and scheme in reference [12] for Audio signal

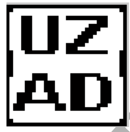
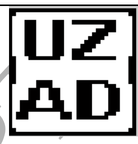
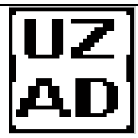
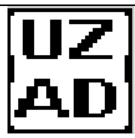
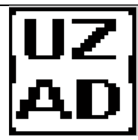
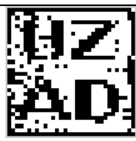
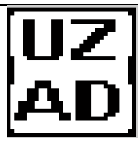
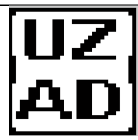
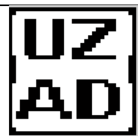
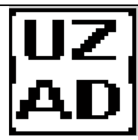
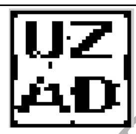
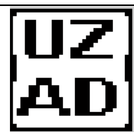
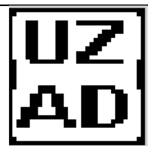
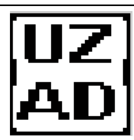
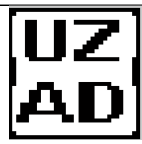
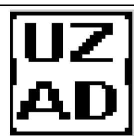
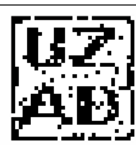
Audio	attacks	Factor (power)	BERs of		NCs of		Detected watermark	
			Scheme in [12]	Proposed scheme	Scheme in [12]	Proposed scheme	Scheme in [12]	Proposed scheme
gspi35_2	AWGN	18 db	00	00	1	1		
	Re-sampling	44100-22050-44100 Hz	00	5.5664	1	0.9558		
	Re-quantization	16-8-16 bits	00	00	1	1		
	Echo	(0.1,0.4)	00	00	1	1		
		(0.3,0.4)	8.6914	00	0.9274	1		
	Amplification	+15%	26.1719	00	0.7591	1		
		-15%	33.4961	00	0.6764	1		
	Cropping	30000	0.8789	00	0.9928	1		
		70000	45.8984	0.0977	0.7447	0.9992		

Table 10: comparisons between our proposed scheme and scheme in reference [13] for Speech signal

Speech	attacks	Factor (power)	BERs of		NCs of		Detected watermark	
			Scheme in [13]	Proposed scheme	Scheme in [13]	Proposed scheme	Scheme in [13]	Proposed scheme
spfe49_1	AWGN	18 db	1.9531	00	0.9841	1		
	Re-sampling	44100-22050-44100 Hz	34.3750	5.1758	0.7026	0.9586		
	Re-quantization	16-8-16 bits	00	00	1	1		
	Echo	(0.1,0.2)	16.6016	00	0.8608	1		
	Amplification	+10%	1.0742	00	0.9913	1		
		-10%	00	00	1	1		
	Cropping	10000	3.5156	00	0.9713	1		
		20000	7.1289	00	0.9414	1		

Authors in [12] and [13] proposed blind watermarking schemes for the audio and speech signals. We compared our proposed design with these published schemes.

Table 9 and Table 10 summarize the comparisons between our proposed watermark detection results and results of schemes in [12] and [13] against various attacks. We observe that the robustness of embedded watermark in our design is better than the embedded watermark in schemes of [12] and [13].

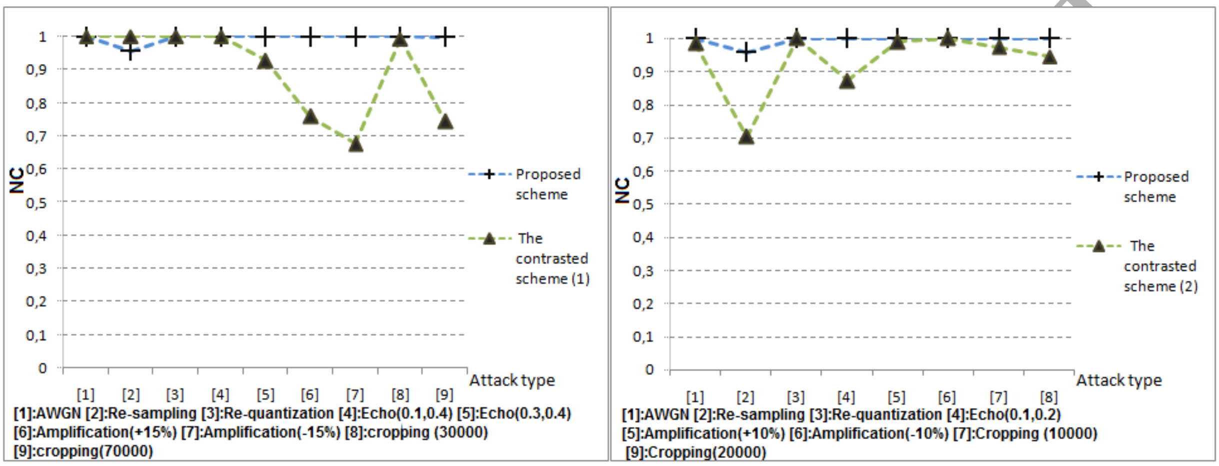


Figure.13: Efficiency comparison between the proposed scheme and other two schemes: the contrasted scheme (1) in [12], and the contrasted scheme (2) in [13]

In Fig.13, the two graphs illustrated well comparison results between our proposed scheme and the two published schemes in references [12] and [13]. Under nine (9) signal processing attacks types, we observe the steady robustness of our proposed design against all strong attacks. Advantages of our proposed design are resumed as:

- It is more robust than the schemes in [12] and [13].
- Our SNR is greater than the SNR determined from scheme of [12] which means better imperceptibility.
- Extraction is blind in our proposed design, without using original signal.
- Extracting without using parameter Δ (the Δ used in the embedding process).
- We can apply both on speech signals and audio signals.

9. Conclusion

In this work, we proposed a new blind scheme for speech and audio signals watermarking based on DWT transformation after framing the original signal and sub-sampling these frames for correlation purpose and applying DCT transform. In order to increase security, Arnold transform is employed. We performed all necessary experiments to ensure the efficiency as well as the fully blind detection is accomplished without using the original speech/audio signal and the insertion parameter is not required. The proposed design, compared to other schemes presented in literatures, makes an excellent tradeoff between security, capacity, imperceptibility and robustness against signal processing attacks at random payload for different types of audio/speech signals. The decomposing with sub-sampling abates a little robustness against the re-sampling attack but gives our proposed design other advantages against other attacks and allows the imperceptibility to remain high.

References

[1] M.A. Akhaee, N.K. Kalantari and F. Marvasti, “Robust audio and speech watermarking using Gaussian and Laplacian modeling”, Signal Processing 90 (2010) 2487–2497, doi:10.1016/j.sigpro.2010.02.013.

[2] H. T. Hu, L.Y. Hsu and H.H. Chou, “Perceptual-based DWPT-DCT framework for selective blind audio watermarking”, Signal Processing (2014),[12pages], http://dx.doi.org/10.1016/j.sigpro.2014.05.003 .

[3] H.T. Hu, L.Y. Hsu and H.H. Chou, “Variable-dimensional vector modulation for perceptual-based DWT blind audio watermarking with adjustable payload capacity”, Digital Signal Processing (2014), http://dx.doi.org/10.1016/j.dsp.2014.04.014 .

[4] H.-T. Hu and L.-Y. Hsu, “ Robust, transparent and high-capacity audio watermarking in DCT domain”, Signal Processing 109 (2015) 226–235, http://dx.doi.org/10.1016/j.sigpro.2014.11.011 .

[5] A. Kaur, M.K. Dutta, K.M. Soni and N. Taneja, “Localized & self adaptive audio watermarking algorithm in the wavelet domain”, Journal of Information Security and Applications 33 (2017) 1-15, http://dx.doi.org/10.1016/j.jisa.2016.12.003 .

[6] B.Y. Lei, I. Y. Soon and Z. Li, “Blind and robust audio watermarking scheme based on SVD–DCT”, Signal Processing 91 (2011) 1973–1984 , doi:10.1016/j.sigpro.2011.03.001.

[7] B.Y. Lei, I.Y. Soon, F. Zhou, Z. Li and H. Lei, “A robust audio watermarking scheme based on lifting wavelet transform and singular value decomposition”, Signal Processing 92 (2012) 1985–2001 doi:10.1016/j.sigpro.2011.12.021 .

[8] H. T. Hu, S. J. Lin and L. Y. Hsu, “Effective blind speech watermarking via adaptive mean modulation and package synchronization in DWT domain”, EURASIP Journal on Audio, Speech, and Music Processing 2017:10, doi:10.1186/s13636-017-0106-4.

[9] H. T. Hu , H.H. Chou, C.Yu and L.Y. Hsu, "Incorporation of perceptually adaptive QIM with singular value decomposition for blind audio watermarking", EURASIP Journal on Advances in Signal Processing, 2014:12, http://asp.eurasipjournals.com/content/2014/1/12 .

[10] C. Yong-mei, G. Wen-qiang and D. Hai-yang, "An Audio Blind Watermarking Scheme Based on DWT-SVD", JOURNAL OF SOFTWARE, VOL. 8, NO. 7, JULY 2013, 1801-1808.

[11] B. Lei, F. Zhou, E. L.Tan, D. Ni, H. Lei, S. Chen and T. Wang, "Optimal and secure audio watermarking scheme based on self-adaptive particle swarm optimization and quaternion wavelet transform", Signal Processing (2014), http://dx.doi.org/10.1016/j.sigpro.2014.11.007 .

[12] X. Wang, P.Wang, P. Zhang, S. Xu, H. Yang, "A norm-space, adaptive, and blind audio watermarking algorithm by discrete wavelet transform", Signal Processing 93 (2013) 913–922, http://dx.doi.org/10.1016/j.sigpro.2012.11.003 .

[13] P. K. Dhar, "A Blind Audio Watermarking Method Based on Lifting Wavelet Transform and QR Decomposition", 8th International Conference on Electrical and Computer Engineering 20-22 December, 2014 , Dhaka, Bangladesh 136-139.

[14] Z. Liu , J. Huang , X. Sun and C. Qi, "A security watermark scheme used for digital speech forensics", Multimedia Tools Appl (2016), DOI 10.1007/s11042-016-3533-9 2016.

[15] M. Ogura, Y. Sugiura and T. Shimamura, "SVD Based Audio Watermarking Using Angle-Quantization", International Conference on Electrical, Computer and Communication Engineering (ECCE), February 16-18, 2017, Cox's Bazar, Bangladesh.

[16] V. Bhat, K. I. Sengupta and A. Das, "An adaptive audio watermarking based on the singular value decomposition in the wavelet domain", Digital Signal Processing 20 (2010) 1547–1558, doi:10.1016/j.dsp.2010.02.006.

[17] A. Benoraira, K. Benmahammed and N. Boucenna, "Blind image watermarking technique based on differential embedding in DWT and DCT domains", EURASIP Journal on Advances in Signal Processing (2015) 2015:55, DOI 10.1186/s13634-015-0239-5.

[18] E. Brannock, M.Weeks, R. Harrison, "Watermarking with Wavelets: Simplicity Leads to Robustness", Computer Science Department Georgia State University, Southeast on, IEEE, pages 587 – 592, 3-6 April 2008.

[19] L. Debnath, Wavelet transform and their application, Department of mathematics university of central Florida Orlanto USA , PINSA –A ,64 no-06 November 1998,p-685-713.

[20] M. A. Osman, N. H. Ali, "Audio Watermarking Based on Wavelet Transform", Applied Mechanics and Materials Vols 229-231 (2012) pp 2784-2788, Trans Tech Publications, Switzerland, doi:10.4028/www.scientific.net/AMM.229-231.2784.

[21] A.E. Villanueva-Luna, A. Jaramillo-Nuñ ez, D. Sanchez-Lucero, C. M. Ortiz-Lima, J. Gabriel Aguilar-Soto, A. Flores-Gil, M. May-Alarcon, "De-Noising Audio Signals Using MATLAB Wavelets Toolbox", Engineering Education and Research Using MATLAB, Dr. Ali Assi (Ed.), ISBN: 978-953-307-656-0, InTech, (2011). http://www.intechopen.com/books/engineering-education-and-researchusing-matlab/de-noising-audio-signals-using-matlab-wavelets-toolbox.

[22] S. M. Deokar and B. Dhaigude, "Blind audio watermarking based on Discrete wavelet and Cosine transform", 2015, International Conference on Industrial Instrumentation and Control (ICIC), College of Engineering Pune, India. May 28-30,2015.

[23] V. Mehta and N. Sharma, "Secure Audio Watermarking based on Haar Wavelet and Discrete Cosine Transform", International Journal of Computer Applications (0975 – 8887), Volume 123 – No.11, August 2015.

[24] A. Tiwari and M. Sharma, "Comparative Evaluation of Semi Fragile Watermarking Algorithms for Image Authentication", Journal of Information Security, Vol. 3 No. 3, 2012, pp. 189-195, DOI: 10.4236/jis.2012.33023.

[25] Y. Lin, W.H. Abdulla, "Audio Watermark: A Comprehensive Foundation Using MATLAB", DOI 10.1007/978-3-319-07974-5. Springer Cham Heidelberg New York Dordrecht London (2015).

[26] P. K. Dhar and T. Shimamura, "Advances in Audio Watermarking Based on Singular Value Decomposition", SPRINGER BRIEFS IN ELECTRICAL AND COMPUTER ENGINEERING (2015), DOI 10.1007/978-3-319-14800-7.

[27] N.V.Lalitha, Ch.Srinivasa Rao and P.V.Y.JayaSree, "DWT-Arnold Transform Based Audio Watermarking", 2013 IEEE Asia Pacific Conference on Postgraduate Research in Microelectronics and Electronics (PrimeAsia).

[28] Subir and Amit M. Joshi, "DWT-DCT based Blind Audio Watermarking using Arnold Scrambling and Cyclic Codes", 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN).

[29] http://sound.media.mit.edu/resources/mpeg4/audio/sqam/

[30] Chu Wai.C, speech coding algorithms: foundation and evolution of standardized coders, April 2004, ISBN: 978-0-471-37312-4. DOI: 10.1002/0471668850. Copyright©2003. John Wiley & Sons, Inc.

[31] A. Al-Haj, A. Mohammad and L. Bata, "DWT–Based Audio Watermarking", The International Arab Journal of Information Technology, Vol. 8, No. 3, July 2011,326-333.

[32] S. Shokri, M. Ismail, N. Zainal and M. Moghaddasi, "Audio-Speech Watermarking Using a Channel Equalizer", Wireless. Pers. Commun (2017), DOI 10.1007/s11277-017-4095-5.

[33] V. Bhat , K. I.Sengupta and A. Das, "An audio watermarking scheme using singular value decomposition and dither-modulation quantization", Multimedia Tools Appl (2010), DOI: 10.1007/s11042-010-0515-1.

[34] M. Hemis, B. Boudraa and T. M. Meksen," New secure and robust audio watermarking algorithm based on QR factorization in wavelet domain", Int. J. Wavelets Multiresolut Inf. Process. 13, 1550020 (2015) , https://doi.org/10.1142/S0219691315500204 .

[35] M.A. Nematollahi , H.G. Rosales, M.A. Akhaee and S.A.A. Al-Haddad, "Robust digital speech watermarking for online speaker recocognition", Hindawi publishing corporation mathematical problems in engineering 2015, Volume 2015, Article ID 372398, http://dx.doi.org/10.1155/2015/372398 .